DOCUMENT RESUME

ED 237 526 TM 830 759

AUTHOR Thompson, R. Brent
TITLE Strategies to Use and to Avoid When Collecting Attitudinal Data on a Delayed Basis.
PUB DATE 1 May 81
NOTE 36p.; Paper presented at the Annual Meeting of the New England Educational Research Organization (Lenox, MA, May 1, 1981).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Analysis of Variance; *Attitude Measures; Course Evaluation; Feedback; Higher Education; *Rating Scales; *Student Attitudes; *Student Evaluation of Teacher Performance; *Test Format; Testing Problems; Test Items
IDENTIFIERS *Delay Effect; Retrospective Studies (Psychology)

ABSTRACT

Sixty-three students in three undergraduate college classes were divided into three groups to investigate the effect of two delayed self-reporting techniques on attitudinal data toward a course and instructor. Nonparametric tests and intercorrelations of item treatment means, together with two-way analyses of variance on item clusters, indicated that the attitudes of students responding to a "then/now" format were significantly more negative than those of other students. It was concluded that delayed self-reporting (on the order of a two-week delay) would not significantly affect the data if investigators avoid use of the "then/now" question format. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

STRATEGIES TO USE AND TO AVOID WHEN COLLECTING ATTITUDINAL DATA ON A DELAYED BASIS

R. Brent Thompson, Ph.D.
Assistant Professor of Education
University of Maine at Orono

ED237526

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. B. Thompson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ABSTRACT

Sixty-three students in three undergraduate college classes were divided into three groups to investigate the effect of two delayed self-reporting techniques on attitudinal data toward a course and instructor. Nonparametric tests and intercorrelations of item treatment means, together with two-way analyses of variance on item clusters, indicated that the attitudes of students responding to a "then/now" format were significantly more negative ($p < .05$) than those of other students. It was concluded that delayed self-reporting (on the order of a two-week delay) would not significantly affect the data if investigators avoid use of the "then/now" question format.

Research Paper presented at the New England Educational Research Organization Annual Meeting in Lenox, Mass., 1 May 1981.

TM 830 715

STRATEGIES TO USE AND TO AVOID WHEN COLLECTING ATTITUDINAL DATA ON A DELAYED BASIS

A frequent activity in educational evaluation is the collecting of attitudinal self-reports (Bloom, Hastings, and Madaus, 1971; Stanley and Hopkins, 1972). At times these data either must be collected on a delayed basis or are more appropriately collected after delay. Delayed measurement of attitudes is more appropriate, for example, when attitude might reasonably be expected to interact with performance on a task. Thus, it could be inappropriate to ask students about their feelings about mathematics, mathematic courses, or their self-confidence in working mathematics at the start of a remedial mathematics course. Calling attention to negative attitudes under these circumstances might be expected to have a similar effect to the questions commonly used by athletes to undermine their opponents' performance (e.g., "Are you still having trouble with your backhand, George?")

Furthermore, as data in a project are gathered, researchers may be reminded of other data which ought to have been collected to help interpret the data now

coming in. Thus they find it necessary to gather the extra data after a delay. While this situation is not desirable, that undesirability alone is not sufficient to avoid the problem in all cases or to correct it in any case. Out of practical necessity, therefore, it is assumed that such data are valid; however, a more empirical basis for this assumption would be preferable.

This study was designed to answer the following question: Do data regarding student attitudes toward a course and instructor at a given point in time, when gathered after a delay (e.g., after two weeks), provide the same information about the course and instructor as would have been gathered at the time?

Two hypotheses related to this question were formulated for this investigation. First, when student attitudes toward a course and instructor at a given point in time are reported on a delayed basis, those delayed reports differ significantly from attitudes reported without delay.

The second hypothesis, which follows, is suggested by a practice commonly encountered by the investigator, that of collecting delayed attitude self-reports with questions using a then/now comparison format (i.e.,

essentially, "How did you feel about . . . x time ago?"/ "How do you feel about . . . now?"). The second hypothesis is that when students are asked to report their attitudes toward a class and instructor at a given point in time after a delay, using a then/now comparison format, the "then" attitudes (i.e., retrospective attitudes about the first point in time) differ significantly from attitudes previously reported at the first point in time.

Method

Treatments

The experiment included three treatments. Students in each of the three treatments reported their attitudes toward the course and the instructor, yielding scores on 36 dependent variables. For each treatment, information was sought concerning students' attitudes at a point approximately midway through the course. Treatments differed on two points: first, whether the data was gathered immediately or after a delay, and second, whether or not those students reporting after a delay were asked to compare their earlier attitudes with their attitudes at the time of reporting.

For students in Treatment 1 (No Delay), the reporting actually took place at midterm. Students in Treatment 2 (Delay) were asked two weeks later to remember

their attitudes toward the course and the instructor at midterm, and to report those attitudes retrospectively.

Students in Treatment 3 (Delay/Compare) were asked two weeks after midterm to report their attitudes about the course and the instructor at the midterm (that is, two weeks previously). In addition, they were asked to report their attitudes about the course and instructor at the "present time" (that is, at the time of reporting). Parallel questions, "How would you have rated . . . two weeks ago?" and "How would you rate . . . now?" were arranged side by side on the form. Thus, course experiences during the two weeks would be allowed to play a more overt role in influencing the attitude reports.

Each of the classes was to have a midterm examination about the time of the experiment. Therefore, the timing of the experiment was chosen so that the midterm examination would occur during the two-week experimental delay period in each class. It was expected that an intervening event such as a midterm would affect attitudes toward the course. Thus, then/now comparisons were expected to yield some differences.

Subjects

Sixty-three students from three undergraduate classes in a large university were the subjects in the present

study. Sixteen of the students were from a horticulture class, 26 from one section of an introductory child development class, and 21 from a second section of the same child development class. The two sections of the child development class were taught by the same individual. Both the child development instructor and the horticulture instructor were very willing to have evaluation questions asked about their courses and themselves.

The total number of potential subjects, based on class attendance when the Treatment 1 data were collected, was 76. However, eight of the students assigned to Treatment 2 and five of the students assigned to Treatment 3 were absent when those data were collected two weeks later. Appendix A contains tables showing the detailed breakdown of attendance at the two sessions of the experiment. Students were not aware that the experiment would include a second session two weeks after the Treatment 1 data collection session.

The self-report forms consisted of 36 items: 22 semantic differential scales (Osgood, Suci, and Tannenbaum, 1957; Kelly, 1963) concerning the general qualities of the course and instructor and certain specific attributes of the instructor; and 12 rating scales (Gronlund,

1971; Stanley and Hopkins, 1972) concerning specific aspects of the course.

Figure 1 shows several examples of the semantic differential scales used in the study. Seven response alternatives were used for these scales, in accordance with findings by Osgood et al (1957) that a greater number of alternatives produced unsatisfactory distributions, while a smaller number of alternatives irritated subjects responding to this type of scale.

Insert Figure 1 about here

Figure 2 shows several examples of the rating scales used in the study. The use of five response alternatives for these scales was dictated by the meanings attached to the responses and follows the common practice of many of the examples cited in Gronlund (1971) and in Stanley and Hopkins (1972).

Insert Figure 2 about here

The items were the same for all treatments. The only changes were in the instructions. ("How would you rate . . . now?" for students in Treatment 1: "How would you have rated . . . two weeks ago?" for students

in Treatment 2; both sets of instructions for students in Treatment 3). All self-report forms also had a cover sheet consisting of demographic questions. This sheet served to mask the differences among the various instruments for the students.

Procedure

The experiment took place during Brigham Young University's Spring term. The Spring term is organized so that a sixteen-week semester is compressed into eight weeks, with classes meeting twice as many hours per week as during a regular semester.

Attitude self-report forms were administered to each class in a group situation. At the time the data were collected for Treatment 1, each student was randomly assigned to one of the three treatments. Students in all treatments were present in each class during the data collection for Treatment 1. Students in Treatment 1 received an attitude report form to complete at that time. At the same time, students in the other treatment groups received an irrelevant form to complete concerning the library on campus. The library questionnaire served two functions: first, to keep the students in Treatments 2 and 3 occupied; and second, to reinforce a reference point for the retrospective questions which they were to answer two weeks later.

This reference point was helpful in the sense that the instructions for Treatment 2 and Treatment 3 on the attitude report form referred to the time "the library survey was given;" the investigator also referred to the time "two weeks ago when I was here before; that is, when you took the library survey."

After the two-week delay, the data were collected from the students in Treatment 2 and Treatment 3. At this time, the students in Treatment 1 also completed the same form that the Treatment 2 students completed.

Data Analysis and Results

Three techniques were used to analyze the data: (1) analyses of variance on clusters of items, (2) a cumulative contingency table based on treatment means for the individual items, and (3) intercorrelations among treatment means on the individual items.

For the analyses of variance, the 36 items were collapsed into five clusters. This was done to reduce the ratio of dependent measures to cell size and to make the results of the analyses more easily interpretable. The five clusters were chosen a priori keeping two guidelines in mind. First, the clusters should provide meaningful aggregates of the individual items. Second, the clusters should be as nearly equal as possible

in the number of items included so that each cluster conveys approximately the same amount of information. When these two criteria came in conflict with each other, however, the first criterion was more closely adhered to. The five clusters thus obtained were: (1) general impression of the course; (2) organization of the course and text; (3) assignments, tests, and grading; (4) instructor benevolence; and (5) instructor competence. Table 1 shows which items were included in each cluster.

Four items from the original 36 were completely eliminated as the clusters were formed. Item 5 (course difficulty: easy--hard) and Item 35 (level of teaching: far too high--far too low) were eliminated because the interpretation of those scales was qualitatively different from all other items. These were the only two items which had as the most favorable response the midpoint of the scale; all other items had one end of the scale as the most favorable response. Two other items (Item 33, instructor interest in the subject; and Item 34, clarity of instructor explanations and instructions) were too ambiguous to be classified as uniquely related to instructor benevolence or competence.

In order to form the clusters, individual scores on each item were transformed to standardized scores.

The mean of the standardized scores in a cluster was then used as the individual's score for that cluster. Separate two-way analyses of variance (with treatment and class as the main effects) were then run for each of the five clusters. Since cell sizes were unequal, the analyses were run using an unweighted-means analysis (Winer, 1971). Table 2 shows the results of those analyses.

Of the five clusters, three exhibited a significant difference due to the treatment main effect ($p < .05$). One cluster exhibited a significant difference due to the class main effect ($p < .05$). There were no significant interactions. The three clusters with significant treatemnt main effects were Cluster 2 (Organization of the course and text), Cluster 3 (Assignments, tests, and grading), and Cluster 5 (Instructor competence). Cluster 3 (Assignments, tests, and grading) was also the cluster with the significant class main effect.

Multiple comparisons of the treatment means were made for Clusters 2, 3, and 5 using Duncan;s New Multiple Range Test. For Cluster 2, the mean response of students in Treatment 3 toward the organization of the course and test was more negative than that of students in Treatment 1. For Cluster 3, the mean response of

---

Insert Table 1 and Table 2 about here

---

students in Treatment 3 toward assignments, tests, and grading was more negative than the mean response of students either in Treatment 1 or Treatment 2. For Cluster 5, no significant differences among individual means emerged.

The second analytic technique used was a cumulative contingency table based on treatment means for the individual items. The treatment means for each item were also categorized as worst (most negative about the course or instructor for the item), middle, or best (most positive about the course or instructor for the item). The cumulative profile of each treatment over all 36 items was analyzed by totaling the number of worst, middle, and best values for each treatment and entering those in a 3 x 3 contingency table (see Table 3). The resulting χ^2 for the table was 46.33 ($p < .05$ with 4 degrees of freedom, C = 0.55). An inspection of Table 3 shows that Treatment 1 means were best (most positive) for 22 of the 36 items, and Treatment 3 means were worst (most negative) for 26 of the 36 items.

Insert Table 3 and Table 4 about here

While the opposite polarities of the distributions of means for Treatment 1 and Treatment 3 were readily apparent, the other specific relationships were not immediately evident. In order to make specific comparisons, Table 2 was collapsed to a 3 x 2 table for analysis using Ryan's procedure (Ryan, 1960; Linton and Gallo, 1975) by deleting the row of "middle" attitude scores. This table (Table 4) had an overall χ^2 of 38.19 with 2 degrees of freedom ($p < .05$); therefore, Ryan's procedure could be applied. The results of the pairwise comparisons indicated that the means of the Treatment 3 student responses were more negative overall toward the course and instructor than the means of either the Treatment 1 or Treatment 2 student responses. No significant difference emerged for the Treatment 1--Treatment 2 comparison.

The third analytic technique used was the intercorrelation of treatment means for 34 of the 36 items. Again Item 5 (course difficulty: easy--hard) and Item 35 (level of teaching: far too high--far too low) were discarded because the most favorable response for these two items was the midpoint of the scale rather than one end of the scale. Correlations were calculated for five treatment conditions: (1) Treatment 1, (2) Treatment 2,

(3) Treatment 3, (4) the Treatment 1 retrospective responses collected at the end of the two-week delay (Treatment 1'), and (5) the Treatment 3 "now" responses (Treatment 3'). Table 5 shows the treatment means for all items for the five treatment conditions. The intercorrelations are shown in Table 6. The two highest intercorrelations were obtained for the Treatment 3--Treatment 3' means (r = .95) and the Treatment 1--Treatment 1' means (r = .87). Closely associated with these intercorrelations are the shifts or distances among the mean item scores for the different treatments. Table 7 shows these mean distances (means of the absolute values of the distances among item means). Again the closest associations (i.e., smallest mean distance) are between Treatment 3 and Treatment 3' (a mean shift or distance of 0.15 points) and between Treatment 1 and Treatment 1' (a mean shift or distance of 0.24 points).

Insert Table 5 through Table 7 about here

The results of the analyses of variance and of the contingency tables indicate that attitudes reported retrospectively after a two-week delay using a then/now comparison format do differ significantly from

attitudes reported without any delay; thus, the second hypothesis of this study is accepted. However, the results do not indicate that the other delayed reporting technique, simple retrospective reporting, produced any significant differences when compared with attitudes reported without delay; thus, the first hypothesis of this study is not accepted.

Discussion

Two logical questions at this point are: (1) Why were the Treatment 3 responses significantly different from the Treatment 1 responses, and (2) Why was there a shift toward more negative attitudes in the Treatment 3 responses? Considering the first question, one possible response is that the students in Treatment 3 felt that an implied demand of the experiment was to find a difference between their "then" and "now" scores. The intercorrelations of means for Treatment 3 ("then" scores) and Treatment 3' ("now" scores) only partially helps resolve that question. The coefficient of .95 indicates that the scores created parallel profiles over the items in the self-report form, but it does not indicate the distance between those profiles. The results of Table 7, however, indicate that the mean distance from "then" to "now" mean scores on items was

only 0.14. Thus it appears unlikely that students in Treatment 3 were trying to create a change situation from "then" to "now." Instead, it appears more likely that both "then" and "now" responses were being answered from the same reference point, "now." This would account for the overall shift as well as the close relationship of "then" and "now" scores.

To see why the overall shift in Treatment 3 attitudes is accounted for by the notion of a single reference point for both "then" and "now" responses, two points need to be made. The first point is that Treatment 2 responses did not differ significantly from the Treatment 1 control scores but did differ significantly from the Treatment 3 "then" scores. The second point is that the Treatment 1' scores (where Treatment 1 students reported restrospectively after the two-week delay) look very much like the original Treatment 1 scores collected without delay ($r_{1,1'} = .87$, mean absolute value of the distance between Treatment 1 and Treatment 1' item means is 0.24). In short, everybody was able to look back at the midterm and report attitudes that were very close to the original nondelayed reports except students in Treatment 3. Thus, Treatment 3 students, confronted with both "then" and "now" questions,

did not seem to be able to overcome the "now" point of view in trying to answer "then" questions.

The second major question posed earlier in the discussion was why the shift was toward more negative attitudes toward the course and instructor in the Treatment 3 "then" scores. One explanation is that the major intervening event in the two-week period was a major examination. If students in Treatment 3 only could not divorce themselves from a "now" point of view, then any dissatisfactions produced by the test would taint their "then" responses. That possibility makes the notion of a single point of view for Treatment 3 students even more attractive because the same notion helps explain both major questions which were posed concerning the shift in Treatment 3 attitudes.

There is, however, a second explanation to account for the shift in Treatment 3 attitudes and the negative direction of that shift. This explanation, a rather pedestrian one it must be admitted, is simply that the Treatment 3 self-report form appears to be twice the burden of the "normal" questionnaire, which only asks about one point in time. While students may not object to a long questionnaire per se, when the length comes from a then/now format the students may well feel that

the investigator is merely trying for double mileage from his questionnaire at their expense.

Further investigation of the attitude shift and its direction could help in deciding which of the above explanations accounts for these phenomena or whether both processes (perceived burden and point of view) are at work. For instance, what would happen if students were divided into three groups, one given only "now" questions, the second given only "then" questions, and the third given the format all in a single administration? If the "perceived burden" explanation were correct, one would expect each set of responses on the then/now instrument to be more negative than the comparable responses from the appropriate comparison group. If, however, the "single point of view" explanation on the then/now instrument to be closely related to the now-only group responses and different from the then-only responses.

Another way of exploring these alternative explanations would be to change the intervening event to a positive one, such as a field trip, a simulation experience, or other positive event. Then using four treatments (before, after using "then" questions, after using "now" questions, and after using a then/now format), one could compare the then/now scores to the various other groups to determine the relative roles of both explanations.

In summary, it is often necessary to collect data related to student attitudes about a course or an instructor. When the data need to be collected after a delay of a few weeks, it appears they may be collected without adversely affecting their validity so long as the then/now comparison format is avoided.

References

Bloom, B. S., Hastings, J. T., and Madaus, G. F. Handbook of formative and summative evaluation of student learning. New York: McGraw-Hill Book Company, 1971.

Gronlund, N. E. Measurement and evaluation in teaching. New York: McMillan Publishing Company, Inc., 1971.

Kelley, G. A. A theory of personality: the psychology of personal constructs. New York: W. W. Norton and Company, 1963.

Linton, M. and Gallo, P. S., Jr. The practical statistician: simplified handbook of statistics. Monterey, California: Brooks/Cole Publishing Company, 1975.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. The measurement of meaning. Urbana, Illinois: University of Illinois Press, 1957.

Ryan, T. A. Significance tests for multiple comparison of proportions, variances and other statistics. Psychological Bulletin, 1960, 57, 318-328.

Stanley, J. D. and Hopkins, K. D. Educational and psychological measurement and evaluation. Englewood Cliffs, New Jersey: Prentice-Hall, 1972.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill Book Company, 1971.

TABLE 1

ITEMS USED IN EACH CLUSTER

Cluster	Item	
	Original Item Number	Description
1. General Impression of the Course	1.	Significant--Trivial
	2.	Useful--Not Useful
	3.	Interesting--Uninteresting
	4.	Important--Unimportant
	6.	Enjoyable--Boring
	7.	Worthwhile--Worthless
	8.	Satisfying--Frustrating
2. Organization of the Course and Text	9.	Clarity of Organization
	10.	Usefulness of Objectives
	11.	Clarity of Objectives
	12.	Usefulness of Text
	13.	Clarity of Text
	14.	Sufficient Examples in Text
3. Assignments, Tests, and Grading	15.	Assignment Instructions are Clear
	16.	Assignments are Significant
	17.	Assignments are Tied in Closely to What is Being Studied in Class
	18.	Tests are Tied in to Instructions

TABLE 1 (continued)

ITEMS USED IN EACH CLUSTER

Cluster	Item: Original Item Number	Item: Description
	19.	Test Questions are Fair
	20.	Feedback is Given Quickly
	21.	Grading Policy is Clearly Explained
	22.	Grading Policy is Fair
4. Instructor Benevolence	23.	Fair--Unfair
	24.	Sincere--Insincere
	26.	Tolerant--Intolerent
	29.	Concerned--Unconcerned
	30.	Available--Unavailable
	36.	The Instructor is Interested in the Students Very Much--Not at All
5. Instructor Competence	24.	Organized--Unorganized
	27.	Confident--Unconfident
	28.	Prepared--Unprepared
	31.	Knowledgeable--Ignorant
	32.	Efficient--Inefficient

TABLE 2

TREATMENT MEANS, CLASS MEANS, AND F-VALUES FOR ALL CLUSTERS

Cluster	Means: Treatment 1. No Delay	2. Delay	3. Delay /Compare	Means: Class 1. Hort	2. CD1	3. CD2	F: Treatment df=2,4	Class $df_{num}=2$	df denom for class
1. General Impression of the Course	9.92	9.88	10.21	9.77	10.06	10.10	1.864	1.406	54
2. Organization of the Course and Text	9.68	10.02	10.28	9.77	9.93	10.17	20.183*	1.536	54
3. Assignments, Tests, and Grading	9.75	9.93	10.34	9.38	10.13	10.28	9.688*	12.126*	54
4. Instructor Benevolence	9.87	10.01	10.14	9.74	10.21	9.94	0.760	2.473	53
5. Instructor Competence	9.73	10.14	10.19	9.68	10.11	10.11	6.988*	2.754	53

Note: For each cluster, a smaller mean indicates a more positive attitude toward the course or instructor.

All scores are based on the standardizing transformation $X' = [\ (X - \bar{X}) / S.D.\ ] + 10$.

* $p < .05$

Table 3

CONTINGENCY TABLE FOR

OCCURRENCE OF TREATMENT MEANS BY ATTITUDE

Attitude	Treatment			Total
	1. No Delay	2. Delay	3. Delay /Compare	
Best	22	13	1	36
Middle	10	17	9	36
Worst	4	6	26	36
Total	36	36	36	108

$\chi^2 = 46.33$ df = 4 $p < .05$

Table 4

COLLAPSED TABLE FOR RYAN'S PROCEDURE

AND RESULTS OF SPECIFIC COMPARISONS

	Treatments			
Attitude	1. No Delay	2. Delay	3. Delay /Compare	Total
Best	22	13	1	36
Worst	4	6	26	36
Total	26	19	27	72

Overall Results

$\chi^2 = 38.19$ $df = 2$ $p < .05$

Specific Comparisons

Treatment 1 vs. Treatment 2	$\chi^2 = 0.86$
Treatment 1 vs. Treatment 3	$\chi^2 = 32.08^*$
Treatment 2 vs. Treatment 3	$\chi^2 = 19.11^*$

$^*p < .05$

TABLE 5

TREATMENT MEANS FOR ALL ITEMS

Item	Treatments				
	1	1′	2	3	3′
1	2.16	2.37	2.11	2.75	3.00
2	1.76	2.32	2.00	2.70	2.70
3	3.72	3.21	2.76	3.30	3.40
4	2.92	2.63	2.83	3.00	3.47
5[a]	4.48	4.47	3.89	3.75	3.60
6	2.60	2.74	3.06	3.75	3.90
7	2.48	2.26	2.28	2.60	2.95
8	3.36	3.00	3.22	3.55	3.75
9	2.16	2.42	2.50	2.80	2.68
10	2.22	2.37	2.76	2.90	2.74
11	2.47	2.47	2.63	3.10	2.95
12	3.00	3.06	2.73	3.28	3.35
13	2.57	2.88	2.94	3.22	3.24
14	2.41	2.53	2.80	2.88	3.19
15	1.80	2.05	1.71	2.40	2.58
16	2.28	2.68	2.00	2.90	3.05
17	2.04	2.53	2.29	2.75	2.63
18	2.43	2.50	2.00	2.60	2.68
19	2.36	2.50	2.44	2.65	2.79
20	2.41	2.17	2.59	2.70	2.68
21	2.44	2.58	3.28	3.45	3.42
22	2.09	2.47	2.67	3.11	3.28
23	2.21	2.21	1.94	2.05	2.15

TABLE 5 (continued)

TREATMENT MEANS FOR ALL ITEMS

Item	Treatments				
	1	1′	2	3	3′
24	1.71	1.95	2.50	2.40	2.40
25	1.67	1.84	2.28	2.10	2.10
26	2.04	2.05	2.22	2.20	2.20
27	1.63	1.79	1.89	1.85	2.05
28	1.50	1.79	2.83	2.55	2.90
29	1.96	2.58	2.17	2.50	2.95
30	2.38	2.79	2.83	2.84	3.05
31	1.42	1.95	1.50	2.25	2.45
32	2.00	2.00	2.28	2.10	2.05
33	1.24	1.68	1.44	1.50	1.40
34	2.28	2.21	2.28	2.60	2.45
35[a]	3.60	3.47	3.61	3.95	4.00
36	2.16	2.42	2.06	2.90	2.80

Note: All means are raw; that is, they have not undergone the standardized score transformation used in the analysis of variance.

[a] Item used in the contingency table, but not in the intercorrelations.

Table 6

INTERCORRELATIONS AMONG TREATMENT MEANS

FOR 34 OF THE 36 ITEMS

	Treatment Comparisons	r
1.	1--1'	.87
2.	1--2	.64
3.	1--3	.75
4.	1--3'	.73
5.	1'--2	.60
6.	1'--3	.84
7.	1'--3'	.82
8.	2--3	.79
9.	2--3'	.75
10.	3--3'	.95

Note. Item 5 and Item 35 were ommitted because the most favorable responses were the scale mid-points rather than one of the scale endpoints.

Table 7

MEAN DISTANCES OF ITEM SCORES BETWEEN PAIRS OF TREATMENTS

	Treatment Comparisons	Means Distance
1	1--1'	0.24
2	1--2	0.34
3	1--3	0.52
4	1--3'	0.60
5	1'--2	0.32
6	1'--3	0.37
7	1'--3'	0.46
8	2--3	0.35
9	2--3'	0.43
10	3--3'	0.15

Note. Distances were obtained by use of the formula $\bar{d} = (|\bar{x}_i - \bar{x}_j|)/N$ where

$\bar{d}$ = the overall mean distance,

$\bar{x}_i$ = the item mean for the first treatment of the comparison,

$\bar{x}_j$ = the item mean for the second treatment of the comparison, and

N = 36, the number of items over which the overall "distance" mean is calculated.

FIGURE 1

SAMPLE SEMANTIC DIFFERENTIAL SCALES

Please indicate how you would rate this COURSE in general (not the instructor) ....

Significant	1	2	3	4	5	6	7	Trivial
Worthwhile	1	2	3	4	5	6	7	Worthless
Satisfying	1	2	3	4	5	6	7	Frustrating

Etc.

Please indicate how you would rate this INSTRUCTOR in general (not the course) ....

Fair	1	2	3	4	5	6	7	Unfair
Concerned	1	2	3	4	5	6	7	Unconcerned
Knowledgeable	1	2	3	4	5	6	7	Ignorant
Efficient	1	2	3	4	5	6	7	Inefficient

Etc.

FIGURE 2

SAMPLE RATING SCALES

Please indicate how you would rate certain specific aspects of the COURSE (not the instructor) ....

	Excellent	Very Good	Good	Fair	Poor
COURSE ORGANIZATION					
Clarity of organization	a	b	c	d	e
Usefulness of objectives	a	b	c	d	e
Clarity of objectives	a	b	c	d	e
TESTS					
Tests are tied in to instruction	a	b	c	d	e
Test questions are fair	a	b	c	d	e
Feedback is given quickly	a	b	c	d	e

ETC.

APPENDIX A

Subject Attendance and Attrition

TABLE A

ATTENDANCE AT THE FIRST EXPERIMENTAL SESSION

Class	Treatments			Totals
	1. No Delay[a]	2. Delay	3. Delay /Compare	
1. Hort	5	7	6	18
2. CD1	12	11	11	34
3. CD2	8	8	8	24
Totals	25	26	25	76

[a] Treatment 1 data came from students in this column.

TABLE B

ATTENDANCE AT THE SECOND EXPERIMENTAL SESSION

Class	Treatments			Totals
	1. No Delay[a]	2. Delay[b]	3. Delay /Compare[c]	
1. Hort	5	6	5	16
2. CD1	9	7	7	23
3. CD2	5	5	8	18
Totals	19	18	20	57

[a] Treatment 1' data came from students in this column.

[b] Treatment 2 data came from students in this column.

[c] Treatment 3 and Treatment 3' data came from students in this column.

TABLE C

SUBJECT ATTRITION

Class	Treatments			Totals
	1. No Delay[a]	2. Delay[b]	3. Delay /Compare[c]	
1. Hort	0	1	1	2
2. CD1	3	4	4	11
3. CD2	3	3	0	6
Totals	6	8	5	19

[a] This column represents the loss of actual subjects from Treatment 1 to Treatment 1'.

[b] This column represents the loss of potential subjects for Treatment 2.

[c] This column represents the loss of potential subjects for Treatment 3 / 3'.